UNITED STATES DISTRICT COURT DISTRICT OF MASSACHUSETTS

Civil Action No. 1:19-cv-12551-FDS
Hon. F. Dennis Saylor IV
ORAL ARGUMENT REQUESTED

Defendant.

PLAINTIFF'S OPPOSITION TO DEFENDANT'S MOTION TO STRIKE EXPERT REPORT OF SUNIL KHATRI, PH.D.

Paul J. Hayes (BBO #227000) Matthew D. Vella (BBO #660171) Kevin Gannon (BBO #640931) Brian Seeve (BBO #670455) Daniel McGonagle (BBO #690084) PRINCE LOBEL TYE LLP

One International Place, Suite 3700 Boston, MA 02110 Tel: (617) 456-8000

Email: phayes@princelobel.com Email: mvella@princelobel.com Email: kgannon@princelobel.com Email: bseeve@princelobel.com Email: dmcgonagle@princelobel.com

ATTORNEYS FOR THE PLAINTIFF

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	BACKGROUND – NO NEW INFRINGEMENT THEORY	2
III.	LEGAL STANDARDS	5
IV.	ARGUMENT	6
	A. PRECLUSION IS A DISFAVORED ONEROUS SANCTION	6
	B. HARMLESSNESS	7
	C. RULE 37(b) FACTORS ALSO WEIGH AGAINST PRECLUSION	8
	1. Justification	8
	2. Ability to Overcome	10
	3. Litigation Abuse	10
	4. Need	11
	5. Docket Impact	11
V.	GOOGLE TECHNICAL MISSTATMENTS	12
VI.	CONCLUSION	12

TABLE OF AUTHORITIES

Cases

Allen v. Martin Surfacing, No. 05-40048-FDS, 2008 WL 6998052 (D. Mass. Sept. 24, 2008)
Brodbeck v. Massachusetts Dep't of Corrections, No. 18-cv-10855, 2021 WL 3131601 (D. Mass. July 23, 2021)
BRT Management LLC v. Malden Storage, LLC, No. No. 17-10005-FDS, 2019 WL 3936862 (D. Mass. Aug. 20, 2019)
D'Pergo Custom Guitars, Inc. v. Sweetwater Sound, Inc., 340 F.R.D. 535 (D.N.H. 2020)
Freeman v. Package Mach. Co., 865 F.2d 1331 (1st Cir. 1988)
Glass Dimensions, Inc. v. State Street Bank & Trust Co., 290 F.R.D. 11 (D. Mass. 2013)passim
Jackson v. Harvard Univ., 900 F.2d 464 (1st Cir. 1990)
Johnson v. H. K. Webster, Inc., 775 F.2d 1 (1st Cir. 1985)11
Lohnes v. Level 3 Commc'ns, Inc., 272 F.3d 49 (1st Cir. 2001)
Mannai Home, LLC v. City of Fall River, No. 17-11915-FDS, 2019 WL 456163 (D. Mass. Feb. 5, 2019)
Poulis-Minott v. Smith, 388 F.3d 354 (1st Cir. 2004)
Samos Imex Corp. v. Nextel Commc'ns, Inc., 194 F.3d 301 (1st Cir. 1999)
Santiago-Diaz v. Laboratorio Clinico y de Referencia del Este, 456 F.3d 272 (1st Cir. 2006)
Thibeault v. Square D Co., 960 F.2d 239 (1st Cir.1992)

Plaintiff, Singular Computing LLC ("Singular"), respectfully submits this opposition to the motion of defendant, Google LLC ("Google"), to strike portions of the Expert Report of Dr. Sunil Khatri ("Motion" or "Mot."). As set forth below Google's motion is factually unfounded and, in reality, a thinly-disguised motion for summary judgment. As a result, the motion should be denied.

I. INTRODUCTION

Google Motion is built upon a fallacy. According to Google, Singular has performed an "about face" via Dr. Khatri's report because:

Singular's infringement contentions did not map the "LPHDR execution unit" element to the so-called "precision-reducer circuits" in the VPU, as Dr. Khatri now does. Instead, Singular's infringement contentions relied *solely* on multipliers in the MXU as satisfying the claim limitation of an LPHDR execution unit.

Mot. at 2. Contrary to Google's assertions, Singular asserted a two-part infringement theory that involved precision-reducer circuits as well as multiplier circuits from the start. Singular's initial infringement contentions, served almost two and a half years ago, explicitly cite Google's own published documents describing multiplication at "reduced bfloat16 precision." Singular cited the same document in its Amended Complaint, filed nearly three years ago. Thus, Google has been on notice of Singular's two-part infringement theory since the case began.

Further, during the Summer of 2022, (over Google's objections) Singular moved for leave to supplement its infringement contentions to provide *additional* specificity and support for its infringement theories, including additional support for the very two-part infringement theory Google now describes as "completely new." At the time, Google opposed (Dkt. No. 363) and tried to prevent Singular's supplementation, apparently wishing to know as little as possible about Singular's infringement theory. The Court granted Singular's motion, and Singular supplemented its infringement contentions to include source code, specifically identifying the

exact "two stage" theory set forth in Dr. Khatri's report. Thus Google's statement that Singular did not disclose Dr. Khatri's "two stage" LPHDR operation is categorically incorrect.

II. BACKGROUND – NO NEW INFRINGEMENT THEORY

Google argues that Dr. Khatri's expert report:

relies on a completely new infringement theory, one that Singular did not previously disclose at any time – not in its preliminary infringement contentions in September 2020, not in its supplemental contentions in August 2022, and not by any other means.

See Mot. at 1. According to Google, this "completely new" theory is a "two-stage" LPHDR operation, namely "rounding" (*i.e.* converting from 32-bit to 16-bit format) and "multiplication". *Id.* at 1-5.

In fact, this theory of infringement was disclosed as early as in the Amended Complaint filed on March 20, 2020 (Dkt. No. 37), and thereafter in Singular's Preliminary Infringement Contentions dated September 4, 2020, and Supplemental Infringement Contentions dated August 10, 2022. For example, paragraph 90(c) of Singular's Amended Complaint states as follows:

Each of those ALUs is adapted to execute a multiplication operation on a value that was converted to a "bfloat16" format after being taken in as input in 32-bit floating point format (FP32 format" or "float32"). The circuitry for taking a float32 input signal, converting it to a bfloat16 value, and then multiplying the value, is hereinafter an "MXU Reduced Precision Multiply Cell. An MXU Reduced Precision Multiply Cell comprises the part of an MXU ALU that performs a multiplication operation, and circuitry for taking a float32 input signal and converting it to a bfloat16 value.

See Dkt. No. 37 at p. 33 (emphasis added). Thus, contrary to Google's argument, the two-part rounding and multiplication operation within the accused TPU devices was described in March 2020.

At the time the Amended Complaint was served, Singular had not been given access to Google's source code or other internal confidential information. Therefore, Singular relied upon Google's publicly-available documents describing how the accused TPU devices operate as

described above. Accordingly, in support of the above statement, Singular cited, *inter alia*, Google documents stating that:

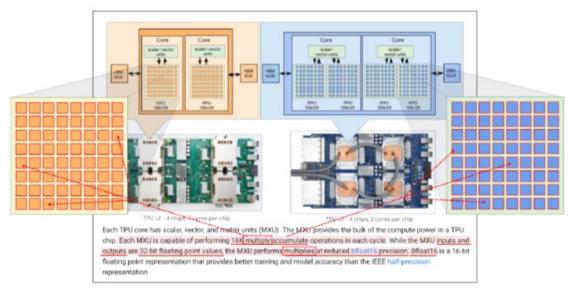
Each TPU core has scalar, vector, and matrix units (MXU). The MXU provides the bulk of the compute power in a TPU chip. Each MXU is capable of performing 16k multiply-accumulate operations in each cycle. While the MXU inputs and outputs are 32-bit floating point values, the MXU performs multiplies at reduced bfloat16 precision.

Amended Complaint, ¶ 90(c) (emphasis added).

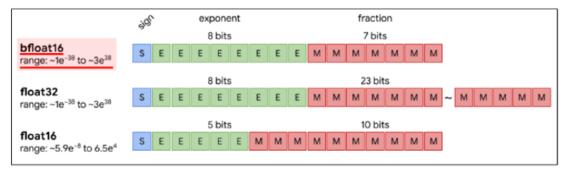
Each MXU takes inputs in FP32 format but then automatically converts them to bfloat16 before calculation.

Id. (emphasis added).

This two-stage rounding/multiplication operation was also described in Singular's Preliminary Infringement Contentions, again using publicly-available Google documents:



https://cloud.google.com/tpu/docs/system-architecture



https://cloud.google.com/tpu/docs/bfloat16

See Ex. A, p. 6 ('156 patent); Ex. B, p. 6 ('273 patent).¹

Google erroneously argues that "Singular's supplemental infringement contentions did not map the 'LPHDR execution unit' element to the so-called 'precision-reducer circuits' in the VPU." Mot. at 1. Twenty two days after receiving copies of the selected portions of the source code Google uses in the accused TPU devices, Singular served its Supplemental Infringement Contentions on August 10, 2022, identifying the printouts (GOOG-SING-SC-###) for the source code used to perform the infringing rounding and multiplying operations. For example:

- "Each TPU core has scalar, vector, and matrix units (MXU). The MXU provides the bulk of the compute power in a TPU chip.
 Each MXU is capable of performing 16K multiply-accumulate operations in each cycle. While the MXU inputs and outputs are
 32-bit floating point values, the MXU performs multiplies at reduced bfloat16 precision. Bfloat16 is a 16-bit floating point representation that provides better training and model accuracy than the IEEE half-precision representation."
 - https://cloud.google.com/tpu/docs/system-architecture
- "The following figure shows three floating-point[] formats
 - fp32 IEEE single-precision floating-point
 - fp16 IEEE half-precision floating point
 bfloat16 16-bit brain floating point
 - https://cloud.google.com/tpu/docs/bfloat16

exponent fraction 8 bits 7 bits bfloat16 M M M M E M M range: ~1e float32 S E E E M M M E Ε float16 S E E E E E M M M M M M M M M M range: ~5.9e-8 to 6.5e

Id

See also GOOG-SING-SC-45-61, 435-444, 449-454.

¹ Unless otherwise noted, all exhibits are attached to the accompanying Declaration of Kevin Gannon.

Notably, pages GOOG-SING-SC-452-54 are titled and cite directly to the very same "precision reducer circuits" (rounders) in the VPU Google claims are "completely new," but in fact are described (as Dr. Khatri's report confirms) in the source code cited in Singular's supplemental contentions. *See*, *e.g.*, Ex. C, p. 8 ('156 patent); *see also* Ex. D, p. 8 ('273 patent).

As Google's counsel informed the Court at the June 2021 hearing in this case, Google's source code is the "truth" regarding how the TPU devices operate. *See* Ex. E at 35:19-20 ("The ground truth is the system Verilog code").² Thus, Google has been aware of Singular's purported "completely new" infringement theory since Singular filed the Amended Complaint (March 2020) and served its Preliminary and Supplemental Infringement Contentions in September 2020 and August 2022, respectively, i.e. long before Dr. Khatri's report was served in December 2022.

III. LEGAL STANDARDS

"Preclusion and negative inference are grave steps, 'by no means an automatic response to a delayed disclosure." *Jackson v. Harvard Univ.*, 900 F.2d 464, 469 (1st Cir. 1990) (citation omitted). As "preclusion is not a strictly mechanical exercise; district courts have some discretion in deciding whether or not to impose that onerous sanction." *Santiago-Diaz v. Laboratorio Clinico y de Referencia del Este*, 456 F.3d 272, 276 (1st Cir. 2006); *see also Brodbeck v. Massachusetts Dep't of Corrections*, No. 18-cv-10855, 2021 WL 3131601, at *3 (D. Mass. July 23, 2021) ("Motions to strike are generally 'disfavored in practice." (citation omitted)). Thus, "Rule 37(c)(1) 'allows the court to admit belatedly proffered expert evidence if the proponent's failure to reveal it was either substantially justified or harmless." *Poulis-Minott*

² "Verilog" is the programming language in which the relevant source code is written.

v. Smith, 388 F.3d 354, 358 (1st Cir. 2004) (emphasis added) (quoting Lohnes v. Level 3 Comme'ns, Inc., 272 F.3d 49, 60 (1st Cir. 2001)).

When considering whether preclusion is warranted, courts should consider: (1) any justification for the late disclosure; (2) the opponent-party's ability to overcome the late disclosure's adverse effects; (3) the party's history of any litigation abuse; (4) the party's need for the evidence; and (5) the impact on the district court's docket. *See*, *e.g.*, *Glass Dimensions*, *Inc. v. State Street Bank & Trust Co.*, 290 F.R.D. 11, 17 (D. Mass. 2013) (citing *Harriman Hancock Cty.*, 627 F.3d 22, 30 (1st Cir. 2013)). Remedial action may be employed to remedy an alleged late disclosure *in lieu* of preclusion. *See*, *e.g.*, *Jackson v. Harvard Univ.*, 900 F.2d at 469.

IV. ARGUMENT

The "weight and value" of Dr. Khatri's opinions are for the jury to assess, "not for [Google's] counsel . . . imperiously to dictate" via a motion to strike. *See Freeman v. Package Mach. Co.*, 865 F.2d 1331, 1341 (1st Cir. 1988). Remedial action should preferably be employed to remedy an alleged late disclosure. *See*, *e.g.*, *Jackson*, 900 F.2d at 469; *see also Moura v. New Prime, Inc.*, No. 4:17-cv-40166, 2020 WL 8570854, at *3 (D. Mass. Dec. 11, 2020) (defendants "had more than three months to prepare for trial from the date they received Dr. Gelfand's expert disclosure, which is an adequate period to prepare any rebuttal evidence or depose Dr. Gelfand, if they so choose"). Accordingly, Google's motion to strike should be denied.

A. PRECLUSION IS A DISFAVORED ONEROUS SANCTION

As stated above, preclusion of evidence is an "onerous sanction." *Santiago–Diaz v Laboratorio Clinico*, 456 F.3d at 276. Accordingly, "preclusion of expert testimony is a grave step, not to be undertaken lightly." *Thibeault v. Square D Co.*, 960 F.2d 239, 247 (1st Cir.1992); *see also Allen v. Martin Surfacing*, No. 05-40048-FDS, 2008 WL 6998052, at *2 (D. Mass. Sept.

24, 2008); *Mannai Home, LLC v. City of Fall River*, No. 17-11915-FDS, 2019 WL 456163, at *6 (D. Mass. Feb. 5, 2019). Preclusion is, therefore, disfavored as akin to an excessive "death penalty" form of relief. *See BRT Management LLC v. Malden Storage, LLC*, No. No. 17-10005-FDS, 2019 WL 3936862, at *3 (D. Mass. Aug. 20, 2019); *see also Brodbeck*, 2021 WL 3131601, at *3 ("Motions to strike are generally 'disfavored in practice" (citation omitted)). Here, as Dr. Khatri's infringement report does not present a new infringement theory and, in any event, there is plenty of time left before trial begins in September 2023, issuing a "death penalty" is not required and would not be fair or just.

B. HARMLESSNESS

"Rule 37(c)(1) 'allows the court to admit belatedly proffered expert evidence if the proponent's failure to reveal it was either substantially justified *or* harmless." *Poulis-Minott v. Smith*, 388 F.3d at 358 (emphasis added) (quoting *Lohnes v. Level 3*, 272 F.3d at 60). Thus, an alleged late disclosure "that occurs well before trial, and with sufficient time to permit the opposing party to adjust its litigation strategy to avoid prejudice, *is harmless*." *D'Pergo Custom Guitars, Inc. v. Sweetwater Sound, Inc.*, 340 F.R.D. 535, 537 (D.N.H. 2020) (emphasis added) (citing *Samos Imex Corp. v. Nextel Commc'ns, Inc.*, 194 F.3d 301, 305 (1st Cir. 1999)).

Trial in this case is not scheduled to begin until September 11, 2023. *See* Dkt. No. 403. Accordingly, as in *D'Pergo*, the case is "well before trial." In *Glass Dimensions*, preclusion was denied by Judge Tauro, *inter alia*, because "trial [was] months away and there [was] ample time to cure any prejudice." *See* 290 F.R.D. at 18. Similarly, in *Moura v. New Prime* 2020 WL 8570854, at *3, Judge Hillman found that defendants "had more than three months to prepare for trial from the date they received Dr. Gelfand's expert disclosure, which is an adequate period to prepare any rebuttal evidence or depose Dr. Gelfand, if they so choose." As a result, the alleged late disclosure in this case, nine months before trial, is harmless. Google will have the

opportunity to question Dr. Khatri regarding his report when they take his deposition. Therefore, the Motion should be denied.

C. RULE 37(b) FACTORS ALSO WEIGH AGAINST PRECLUSION

As set forth in Section II above, the following factors are generally considered when addressing a motion to preclude: (1) any justification for the late disclosure; (2) the opponent-party's ability to overcome the late disclosure's adverse effects; (3) the party's history of any litigation abuse; (4) the party's need for the evidence; and (5) the impact on the district court's docket. *See*, *e.g.*, *Glass Dimensions*, 290 F.R.D. at 17. These factors weigh against preclusion here.

1. Justification

Dr. Khatri's report does not introduce a new two stage theory of infringement. Thus, justification is unnecessary. For the sake of completeness, however, Singular will address this factor. Singular has set forth above how it has mapped the rounding and multiplication operations to the accused TPU devices since the beginning of the case in the Amended Complaint and, thereafter, in its original and supplemental infringement contentions. Thus, Google's argument that Singular's infringement theories are "completely new" (Mot. at 1) is without merit.

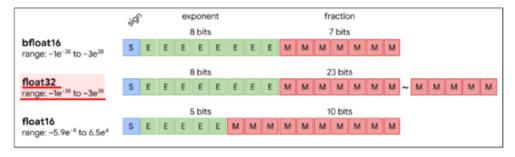
Per Google's limitations on source code production in the Protective Order, Singular was only allowed to request printouts of small percentage (up to 500 pages out of tens of thousands) of the entire TPU source code, and was further limited in the number of consecutive pages (up to 25) of source code files it was allowed to print. *See* Dkt. No. 87 at 17. Google refused to provide printouts of the requested files before the Court stayed the case pending IPR in August 2021. *See* Ex. F ("Given that the case is now stayed following the August 27, 2021 status conference with Judge Saylor, Singular's request for printing of source code at this time is premature"). Only

after the stay was lifted in May 2022 did Google finally (on July 19, 2022) provide printouts of the source code that Singular had requested.

Less than a month after it received printed copies of the relevant source code (on August 10, 2022), Singular served its supplemental contentions³ that included citations to the relevant source code for each and every limitation of the asserted claims. *See* Exs. C-D. For example, Singular highlighted source code citations in yellow in numerous places in its supplemental contentions:

- "Each TPU core has scalar, vector, and matrix units (MXU). The MXU provides the bulk of the compute power in a TPU chip. Each MXU is capable of performing 16K multiply-accumulate operations in each cycle. While the MXU inputs and outputs are 32-bit floating point values, the MXU performs multiplies at reduced bfloat16 precision. Bfloat16 is a 16-bit floating point representation that provides better training and model accuracy than the IEEE half-precision representation."
 https://cloud.google.com/tpu/docs/system-architecture
- · "The following figure shows three floating-point[] formats
 - fp32 IEEE single-precision floating-point
 - · fp16 IEEE half-precision floating point
 - · bfloat16 16-bit brain floating point"

https://cloud.google.com/tpu/docs/bfloat16



Id.

See also GOOG-SING-SC-45-61, 435-444, 449-454

See Ex. C, p. 8 ('156 patent); see also Ex. D, p. 8 ('273 patent). For example, the file

is found on cited pages GOOG-SING-SC-000452-454. *See*Declaration of Kevin Gannon, ¶ 8. As apparent from its title, this code corresponds to the very same "precision-reducer circuits" (i.e. rounders) that Google erroneously claims are "completely new" and first disclosed in Singular's expert report. *See also id.* at ¶ 9 (source code cited in

³ Over Google's objections, the Court granted Singular's motion for leave to supplement its infringement contentions to add citations to Google's source code. *See* Dkt. No. 372.

Singular's supplemental contentions stating that

By providing citations, *inter alia*, to this code, the same code as is cited in Dr. Khatri's report (*see*, *e.g.*, Khatri Report, p. 29, ¶¶ 0144-45),⁴ for each and every limitation of the asserted claims (*see* Exs. C-D), Singular's contentions provided more than adequate notice to Google of the infringement theories set forth in the report.

2. Ability to Overcome

Google has plenty of time within which to present a response to the various paragraphs of the Khatri report that is the subject of the present Motion. Trial is not until September 2023. Dkt. No. 403. In *Glass Dimensions*, preclusion was denied by this Court, inter *alia*, because "trial [was] months away and there [was] ample time to cure any prejudice." *See* 290 F.R.D. at 18. Trial is likewise "months away" in this case. Similarly, in *Moura*, preclusion was denied because "defendants ha[d] substantial time to prepare to confront [plaintiff's expert] at trial." *Moura v. New Prime*, 2020 WL 8570854, at **3-4 (defendants "had more than three months to prepare for trial from the date they received Dr. Gelfand's expert disclosure" which was "an adequate period to prepare any rebuttal evidence or depose Dr. Gelfand, if they so choose"). Expert discovery does not close until March 24, 2023. *See* Dkt. No. 422. Moreover, the material sought to be precluded relates to Google's own code which it wrote. Thus, Google does not have to take any additional discovery understand how its own accused products work in order to prepare a response to the objected-to paragraphs in Dr. Khatri's report.

3. <u>Litigation Abuse</u>

There is no history of litigation abuse by Singular in this case.

⁴ Google filed a copy of Dr. Khatri's report under seal at Docket No. 410-1.

4. Need

In *Glass Dimensions*, preclusion was denied, *inter alia*, because the defendants had "a need for the data in order to demonstrate compliance with [a] defense" that it was asserting. *See* 290 F.R.D. at 18; *see also Moura v. New Prime*, 2020 WL 8570854, at *3 (evidence needed to support expert's opinion). Here, the evidence and testimony of Dr. Khatri that Google would have stricken is needed to assist in proving Singular's claim of infringement. Thus, as in *Glass Dimensions* and *Moura*, the evidence is needed at trial.

5. Docket Impact

Denying the motion to preclude will not adversely affect the docket in this case. As stated, expert discovery does not close until March 24, 2023. *See* Dkt. No. 422. Trial does not commence until September 11, 2023. *See* Dkt. No. 403. Thus, no continuance of either of these dates will be required. *See*, *e.g.*, *Moura v. New Prime*, 2020 WL 8570854, at *3 (three months before trial "is an adequate period to prepare any rebuttal evidence or depose Dr. Gelfand, if they so choose.").

Tellingly, Google does not request any additional discovery, or additional time for expert discovery. This suggests that its Motion may be tactical and, in any event, weighs against preclusion. *See*, *e.g.*, *Johnson v. H. K. Webster*, *Inc.*, 775 F.2d 1, 8 (1st Cir. 1985) ("Courts have looked with disfavor upon parties who claim surprise and prejudice but who do not ask for a [continuance] so they may attempt to counter the opponent's testimony"); *see also Allen v. Martin Surfacing*, 2008 WL 6998052, at *2 (denying motion to strike when defendant did not "seek any sanction other than the 'grave' and 'onerous' one of wholesale exclusion"); *D'Pergo*, 340 F.R.D. at 538 ("The suggestion that D'Pergo will suffer harm absent such exclusion is unpersuasive given D'Pergo's failure to seek any less extreme remedy") (citing *Johnson*, 775 F.2d at 8).

V. GOOGLE TECHNICAL MISSTATMENTS

Google makes at least the following demonstratively incorrect technical statements.

- 1. Google states that Singular's Preliminary Infringement Contentions relied "solely" on multipliers in the MXU. Mot. at 2. As explained above, Singular's Amended Complaint and Preliminary Infringement Contentions, however, both map the LPHDR execution unit to a precision reducer circuit (rounder). *See* Amended Complaint (Dkt. No. 37), ¶ 90(c); Preliminary Infringement Contentions Claim Charts (Ex. A., pp. 6-7; Ex. B, pp. 6-7).
- 2. Google states Singular did not disclose a "two stage" VPU LPHDR operation in its Supplemental Infringement Contentions. Mot at 1. As explained above, Google's "two stage" source code was identified by Singular in its Supplemental Infringement Contentions that specifically map the asserted claims to the precision reducer circuit (rounder) in the VPU. *See supra* (citing Google rounder source code GOOG-SING-SC-000452-454; *see also* Ex. C, pp. 7-8; Ex. D, pp. 7-8).

VI. CONCLUSION

For the reasons set forth above, Google's motion should be denied.

REQUEST FOR ORAL ARGUMENT

Singular believes that oral argument may be of assistance to the Court in resolving the present motion. Accordingly Singular wishes to be heard and requests oral argument pursuant to Local Rule 7(d).

Dated: February 1, 2023. Respectfully submitted,

/s/ Paul J. Hayes

Paul J. Hayes (BBO #227000) Matthew D. Vella (BBO #660171) Kevin Gannon (BBO #640931) Brian Seeve (BBO #670455) Daniel McGonagle (BBO #690084)

PRINCE LOBEL TYE LLP

One International Place, Suite 3700

Boston, MA 02110 Tel: (617) 456-8000

Email: phayes@princelobel.com Email: mvella@princelobel.com Email: kgannon@princelobel.com Email: bseeve@princelobel.com Email: dmcgonagle@princelobel.com

ATTORNEYS FOR THE PLAINTIFF

CERTIFICATE OF SERVICE

I certify that, on February 1, 2023, all counsel of record who have consented to electronic service are being served with a copy of this document via the Court's CM/ECF system.

/s/ Paul J. Hayes